# Rare Variant Alleles in the Light of the Neutral Theory[1]

*Motoo Kimura*
National Institute of Genetics



Based on the neutral theory of molecular evolution and polymorphism, and particularly assuming "the model of infinite alleles," a method is proposed which enables us to estimate the fraction of selectively neutral alleles (denoted by $P_{neut}$) among newly arisen mutations. It makes use of data on the distribution of rare variant alleles in large samples together with information on the average heterozygosity. The formula proposed is $P_{neut} = [\bar{H}_e/(1 - \bar{H}_e)]\ [\log_e(2\bar{n}q)/\bar{n}_a(x < q)]$, where $\bar{n}_a(x < q)$ is the average number of rare alleles per locus whose frequency, $x$, is less than $q$; $\bar{n}$ is the average sample size used to count rare alleles; $\bar{H}_e$ is the average heterozygosity per locus; and $q$ is a small preassigned number such as $q = 0.01$. The method was applied to observations on enzyme and other protein loci in plaice, humans (European and Amerindian), Japanese monkeys, and fruit flies. Estimates obtained for them range from 0.064 to 0.21 with the mean and standard error $P_{neut} = 0.14 \pm 0.06$. It was pointed out that these estimates are consistent with the corresponding estimate $P_{neut}(Hb) = 0.14$ obtained independently based on the neutral theory and using data on the evolutionary rate of nucleotide substitutions in globin pseudogenes together with those in the normal globins.

## Introduction

During the past decade and a half, much attention has been paid to protein polymorphisms (and more recently, DNA polymorphism), and various statistical methods have been developed to analyze the data (see Kimura [1983], pp. 275–281, for review). There has also been much discussion, particularly in the form of the neutralist-selectionist controversy (Crow 1972, 1981; Calder 1973; Lewontin 1974; Harris 1976; Ruffié 1976; Selander 1976), regarding the mechanism by which molecular polymorphisms are maintained.

However, very little attention has been paid to rare variant alleles whose frequencies in the population are too low for them to be regarded as members of polymorphic systems. This is understandable, because such alleles do not make

1. Contribution no. 1488 from the National Institute of Genetics, Mishima, Shizuoka-ken, 411 Japan. Key words: protein polymorphism, population genetics, molecular evolution, neutral mutations.

Address for correspondence and reprints: Motoo Kimura, National Institute of Genetics, Yata 1,111 Mishima, 411, Japan.

*Mol. Biol. Evol.* 1(1):84–93.

0737-4038/83/0101-0003$02.00

any significant contribution to the genetic variability of the species, and also because they cannot be detected unless the sample size is unusually large.

In this paper I intend to show that such rare variant alleles can, nevertheless, supply valuable information on the mechanism by which polymorphism at the molecular level is maintained. In particular, I shall demonstrate that the neutral theory of protein polymorphism (Kimura 1968*a*, 1968*b*; Kimura and Ohta 1971; Kimura [1983] for review) can supply a theoretical basis to connect observations on rare variant alleles with those of polymorphic alleles. Furthermore, I shall endeavor to show that the result of data analysis fits well the larger picture of molecular evolution as seen from the standpoint of the neutral theory.

## Basic Theory

Let us assume a random mating, diploid population of effective size $N_e$. Consider a particular locus and assume the infinite allele model (Kimura and Crow 1964), that is, assume that whenever mutation occurs it leads to an allele not already existing. Let $v$ be the mutation rate per locus per generation. I denote by $\Phi(x)$ the distribution of allelic frequencies under the assumption of an equilibrium in which mutational production of new alleles is balanced by random extinction of existing alleles. This distribution means that $\Phi(x)dx$ represents the expected number of alleles whose frequencies lie in the range between $x$ and $x + dx$. When all of the mutations are selectively neutral, it was shown by Kimura and Crow (1964) that

$$\Phi(x) = 4N_e v(1-x)^{4N_e v-1}x^{-1}. \tag{1}$$

Note that this distribution refers only to those alleles actually contained in the population ($x > 0$); although I assume there are an infinite number of possible alleles, only a limited number of them are present at any given moment in the population, and I do not include countless missing alleles. In the following, I shall use the letter $M$ to stand for $4N_e v$, so that the right-hand side of equation (1) becomes $M(1-x)^{M-1}x^{-1}$.

The average value of the sum of squares of allelic frequencies or the average homozygosity is

$$\bar{H}_0 = \int_0^1 x^2\Phi(x)dx = 1/(M+1), \tag{2}$$

and, therefore, the average heterozygosity is

$$\bar{H}_e = 1 - \bar{H}_0 = M/(M+1). \tag{3}$$

This means that, if I know the value of $\bar{H}_e$ from observation, I can estimate the parameter $M$ by the formula

$$M = \bar{H}_e/(1-\bar{H}_e). \tag{4}$$

As emphasized by Nei (1975), a reliable estimate of the average heterozygosity of any species can only be obtained by averaging heterozygosities over a number of loci. Therefore, it is necessary to consider the possibility that mutation rates for neutral alleles differ among loci. To take such a variation into account, Nei et al. (1976) proposed an infinite allele model assuming that $M$ or $4N_e v$ among loci

follows a gamma distribution with the mean $\bar{M}$ and the variance $V_M$. In this model, the relationship between the average heterozygosity and $\bar{M}$ is more complicated, but Nei (1975) derived a useful approximation formula,

$$\bar{H}_e = \frac{\bar{M}}{1 + \bar{M}} - \frac{V_M}{(1 + \bar{M})^3}, \tag{5}$$

which is valid unless $\alpha = \bar{M}^2/V_M$ is small and $\bar{M}$ is large. According to Nei et al. (1976), an appropriate value of $\alpha$ is about 1. Also, for a wide range of organisms, $\bar{M}$ seldom exceeds 0.3 (see, e.g., Nevo 1978), so this formula should have wide applicability. Note that under these circumstances $V_M$ is much smaller than $\bar{M}$, and therefore variation of $M$ among loci has a relatively small effect on $\bar{H}_e$, as easily seen by comparing equations (3) and (5).

One important point which I should note in estimating $M$ from $\bar{H}_e$ using equation (4) is that $M$ is determined largely by polymorphic alleles; rare alleles contribute very little to $\bar{H}_e$ and therefore to $M$.

Next, let us examine the occurrence of rare alleles whose frequencies are less than a certain small value $q$. Bearing in mind the standard practice of defining a polymorphic locus as one in which the most frequent allele does not exceed 99%, I find it appropriate to take $q = 0.01$.

It can be shown mathematically (see, e.g., Kimura 1983, p. 227) that, in the neighborhood of $x = 0$, the population behavior of alleles in general, including those having mild selective advantage or disadvantage, is essentially the same as that of selectively neutral mutants. Thus the average number of alleles per locus whose frequencies are less than $q$ is

$$\bar{n}_a(x<q) = \int_{1/(2n)}^{q} \Phi(x)dx \approx M \log_e(2nq), \tag{6}$$

where $n$ is the sample size. This formula is valid if $4N_e|s|q$ is small, where $|s|$ is the absolute value of the selection coefficient of a mutant allele. In this formula $M$ stands for $4N_e v$; however, $v$ here represents the mutation rate for practically all types of alleles, as pointed out by Nei (1977), and not just for selectively neutral alleles. In fact, he proposed the use of equation (6) for estimating the mutation rate for protein loci. I shall denote $M$ in this equation by $M_q$ in order to distinguish it from $M$ obtained by equation (4). If rare variants are scored at more than one locus with a large sample for each locus (the mean size being $\bar{n}$ per locus), and if $N_e$ is known, I can estimate the total mutation rate per locus by

$$v_{T(E)} = M_q/(4N_e), \tag{7}$$

where the subscript $E$ refers to electrophoretically detectable alleles, and

$$M_q = \bar{n}_a(x < q)/\log_e(2\bar{n}q). \tag{8}$$

However, if I denote by $v_{0(E)}$ the mutation rate per locus for selectively neutral (and electrophoretically detectable) alleles, then $M = 4N_e v_{0(E)}$, so that

$$v_{0(E)} = M/(4N_e). \tag{9}$$

Thus, I can estimate the fraction of neutral alleles among all the mutations that can be detected electrophoretically by

$$P_{\text{neut}} = v_{0(E)}/v_{T(E)} = M/M_q. \tag{10}$$

What is important here is that, even when the actual value of $N_e$ is not known, we can estimate the fraction of neutral alleles at the time of their occurrence by the ratio $M/M_q$, provided that the neutral theory is correct. This equation may be rewritten as

$$P_{\text{neut}} = \frac{\bar{H}_e}{1-\bar{H}_e} \cdot \frac{\log_e(2\bar{n}q)}{\bar{n}_a(x<q)}, \tag{10a}$$

where $\bar{H}_e$ is the mean heterozygosity per locus estimated by averaging over a number of polymorphic as well as monomorphic loci, $\bar{n}_a(x < q)$ is the number of rare variant alleles per locus, and $\bar{n}$ is the average sample size over loci used to count the rare alleles whose frequencies are less than $q$. As mentioned already, an appropriate value for $q$ is 0.01, although other values, such as $q = 0.005$, may be used.

Equations (6)–(10a) contain several assumptions and approximations. In particular, equation (6) is derived by assuming that the distribution of rare alleles in the sample is sufficiently close to that in the population. More accurate (but more complicated) formulas on the subject have been derived by Chakraborty (1981). Also, the use of the average sample size $\bar{n}$ in equation (8) requires that variation of the sample size among loci is relatively small. In the plaice data, $n$ varies around the mean = 1,956 with the standard deviation 508, so that use of the mean ($\bar{n}$ = 1,956) alone will not cause much error. It is hoped that these approximations are acceptable for the moment, and that, in the future, more extensive data will be analyzed with better statistical methods.

## Data Analysis

As the first example of the application of the theory above to estimate the fraction of selectively neutral mutations, I shall use the data from Ward and Beardmore (1977) on protein variation in the plaice, *Pleuronectes platessa*, a marine flatfish. They screened electrophoretically detectable variation at 46 protein loci (39 enzyme and seven nonenzyme proteins), taking very large samples from the Bristol Channel population. This probably represents the most comprehensive investigation of protein variation in fishes. The mean heterozygosity per individual per locus turned out to be 0.102 ± 0.026. The sample sizes differ from locus to locus: they are more than 2,000 in 8 loci, between 2,000 and 500 in 9 loci, between 500 and 100 in 16 loci, but less than 100 in the remaining 13 loci. Using equation (4), I get $M = 0.114$ for $\bar{H}_e = 0.102$. If I use Nei's formula 5, then $\bar{M} = 0.128$ for $\alpha = 1$ and $\bar{H}_e = 0.102$, so the effect of variation of mutation rate among loci is rather small. Therefore, in the following, I shall use equation (4) to simplify our calculation.

In order to estimate $M_q$ from observations on rare variants, the sample size must be large. Therefore, I have chosen from the data of Ward and Beardmore (1977, tables 2 and 3) 11 loci for which the sample size per locus is larger than 1,000. The average sample size per locus for them turned out to be $\bar{n}$ = 1,956 ±

508. Of these 11 loci, 8 are polymorphic and 3 are monomorphic. The average heterozygosity of these 11 loci is 0.147, which is not very different from the average heterozygosity of 46 loci, that is, 0.102. Among these loci, 30 alleles are found whose frequencies in the sample are less than 0.01. Thus, $\bar{n}_a(x < 0.01) = 30/11 = 2.73$ per locus. Then, applying equation (8), where I assume $\bar{q} = 0.01$, I get $M_q = 2.73/\log_e(2 \times 19.56) = 0.744$. This leads to $v_{T(E)} = 0.744/(4N_e)$. Although the real value of $N_e$ is not known, if it is $10^6$, we have $v_{T(E)} = 1.86 \times 10^{-7}$. Finally, substituting $M = 0.114$ and $M_q = 0.744$ in equation (10), I get $P_{neut} = 0.15$. This means that one mutation out of 6.5 on the average is selectively neutral while the remaining 5.5 are too deleterious to contribute to protein polymorphism.

As the second example, I shall use the data on human populations of Harris et al. (1974), who reported the incidence of rare alleles determining electrophoretic variants at 43 enzyme loci in Europeans. From their table 1, I have chosen 26 loci for which the sample size is larger than 1,000. The average sample size for them is $\bar{n} = 4{,}058.04$. The average number of rare alleles per locus has turned out to be 49/26 or 1.88. Since Harris et al. defined rare alleles as those alleles whose individual frequency in the sample was less than 0.005, this corresponds to $q = 0.005$ of equation (8). Then, substituting $\bar{n}_a(x < q) = 1.88$, $\bar{n} = 4{,}058.04$, and $q = 0.005$ in this equation, I get $M_q = 0.508$. This gives an estimate for $4N_e v_{T(E)}$, where $v_{T(E)}$ is the total mutation rate per enzyme locus for electrophoretically detectable alleles. However, from equation (4), I obtain $M = 0.0718$ by assuming $\bar{H}_e = 0.067$, which is an approximate value for the average heterozygosity per locus due to common polymorphic alleles (Harris and Hopkinson 1972). Then from equation (10) the fraction of mutations that are selectively neutral among all electrophoretically detectable mutations is $P_{neut} = M/M_q = 0.14$. This value is very close to the corresponding estimate obtained for the plaice.

Extensive studies of rare variants in human populations have also been done by Neel and his associates on Amerindians, and valuable data have been obtained. I use the data presented in table 1 of Neel (1978), which lists the occurrence of rare variants at 28 loci in 21 Amerindian tribes. His definition of rare variant alleles corresponds to $q = 0.01$ in my terminology, and from his table I obtain $\bar{n}_a(x < q) = 1.29$ and $\bar{n} = 6{,}442.07$, giving $M_q = 0.266$. An interesting feature of his data is that some of the variants represent what he calls "private polymorphisms," that is, they are concentrated in a single or several related tribes where their frequencies are well above the minimum for a polymorphism. For example, an allele called YAN-2 at the albumin (Alb) locus is present in more than 6% of the members of the Yanomama tribe but absent in other tribes. We can calculate the value of $M$ using the average heterozygosity at 23 loci over 12 tribes as listed in table 5 of Neel (1978), where I find $\bar{H}_e = 0.054$. Thus, I obtain $M = 0.057$. Therefore, the fraction of neutral mutations, as estimated by $M/M_q$, turns out to be $P_{neut} = 0.21$, which is not very different from the corresponding value obtained for European populations (i.e., $P_{neut} = 0.14$).

The occurrence of rare variants is also reported in the Japanese macaque (*Macaca fuscata fuscata*) studied by Nozawa and his associates (see, e.g., Nozawa et al. 1982). They surveyed 32 independent protein loci and obtained 1.3% as the average heterozygosity, which is a rather low value even for mammals. Their extensive studies so far yield the following data (Nozawa, personal communication, 1981). The average number of rare variants per locus is $\bar{n}_a(x < 0.01) = 23/32 = 0.719$, and the average sample size is $\bar{n} = 1{,}609.9$. Thus I get $M_q =$

0.207 from equation (8). The observed average heterozygosity per locus is $\bar{H}_e$ = 0.013 ± 0.0014, from which I get $M$ = 0.0132. Using these values, I obtain $P_{neut}$ = $M/M_q$ = 0.064. This means that, roughly speaking, one mutation out of 16, on the average, is selectively neutral in Japanese monkeys. This is less than half as large as the corresponding value obtained for the plaice.

The census number of the total population of the Japanese macaque is estimated to be 20,000–70,000. It is also estimated that the effective population size is about one-third of its census number (cited from Nozawa et al. 1975). Following Nei (1977), if I assume $N_e = 2 \times 10^4$, I obtain $v_{T(E)} = M_q/(4N_e) = 2.6 \times 10^{-6}$. However, the mutation rate for neutral alleles is $v_{0(E)} = M/(4N_e) = 1.65 \times 10^{-7}$ per generation.

As the final example, I shall analyze the data from *Drosophila willistoni* group studied by Ayala and his associates (1974). The sample size per locus per species in this case is not as large as in the previous examples. Of the five species studied, only the *D. willistoni* data are extensive enough for the average sample size per locus to be larger than 500 (in terms of the gene number, i.e., $2n$), so I shall concentrate on this species. From their table 1, which lists allelic frequencies at 31 loci, I have chosen alleles whose frequencies are less than 1% ($q$ = 0.01). There are 85 such alleles, so that $\bar{n}_a(x < q) = 85/31 = 2.74$. The average sample size per locus is $2\bar{n}$ = 568.06. From these values, I obtain $M_q$ = 1.60. The average heterozygosity per locus ($\bar{H}_e$) as listed in table 6 of Ayala et al. (1974) is 0.177, from which I obtain $M$ = 0.215. Therefore, the estimate for the fraction of neutral mutations among all electrophoretic mutations at the time of occurrence is $P_{neut}$ = $M/M_q$ = 0.13. This estimate is not very different from the corresponding estimates obtained for human populations, as well as for the plaice.

The results of analyses of the five examples above are summarized in table 1. The average of five $P_{neut}$ values turns out to be 0.14 ± 0.06.

## Discussion

From the standpoint of the neutral theory, the rare variant alleles are simply those alleles whose frequencies within a species happen to be in a low-frequency range (0, $q$), whereas polymorphic alleles are those whose frequencies happen to be in the higher-frequency range ($q$, 1 − $q$), where I arbitrarily take $q$ = 0.01. Both represent a phase of molecular evolution.

**Table 1**
**Proportion of Selectively Neutral Mutations at the Time of Occurrence among Electrophoretically Detectable Mutations ($P_{neut}$) Estimated from Five Data Sets**

| Organism | $\bar{H}_e$ | $M$ | $M_q$ | $P_{neut}$ |
|---|---|---|---|---|
| Plaice (*Pleuronectes platessa*) | .102 | .114 | .744 | .15 |
| Human: | | | | |
| European | .067 | .072 | .509 | .14 |
| Amerindian | .054 | .057 | .266 | .21 |
| Japanese macaque | .013 | .013 | .207 | .064 |
| Fruit fly (*Drosophila willistoni*) | .177 | .215 | 1.60 | .13 |

However, in contrast to polymorphic alleles, which are predominantly neutral, the rare variant alleles may include slightly deleterious and sometimes even definitely deleterious alleles in addition to selectively neutral ones. This means that rare variant alleles reflect the total mutation rate much more faithfully than polymorphic alleles. In this connection, Harris et al.'s (1974) observation is relevant. They found that "polymorphic" and "monomorphic" loci do not differ in the average heterozygosity for rare alleles if the placental alkaline phosphatase, an unusually variable locus, is excluded. This is easy to understand if we note that the intrinsic mutation rates ($v_{T(E)}$) at these two classes of loci may essentially be the same.

In the analysis above, I have estimated the fraction ($P_{neut}$) of selectively neutral mutations at the time of occurrence among mutations that can be detected by electrophoretic method, using data from the plaice, humans, the Japanese macaque, and the fruit fly (see table 1). It is remarkable that this fraction is relatively uniform among widely separated species with highly different average heterozygosities.

The present analysis is consistent with Ohta's (1975) finding on the excess of rare alleles: using *Drosophila* and human data, she noticed that the observed and theoretical distributions of allelic frequencies agree quite well under the neutral theory with respect to polymorphic alleles but that there is a marked excess of rare alleles in the observed distribution. Ohta (1976) went further and showed, using data on *D. willistoni,* that the excess of rare alleles is more pronounced in the substrate-specific enzymes than the substrate-nonspecific enzymes (see Ohta's [1976] table 3). Actually, if we apply the present method for each of these classes of enzymes separately (16 specific and 15 nonspecific enzymes in *D. willistoni*), we obtain $P_{neut} = 0.070$ for the substrate-specific group and $P_{neut} = 0.204$ for the substrate-nonspecific group. From these two $P_{neut}$ values, it is evident that the probability of a mutational change being selectively neutral is much smaller for the substrate-specific enzymes than for the nonspecific enzymes.

A similar calculation can be done using human data (Harris et al. 1974) following the classification of group I (substrate-specific) and group II (substrate-nonspecific) enzymes proposed by Gillespie and Langley (1974). For 13 loci of the group I enzymes, it turns out that $P_{neut} = 0.11$, and for 10 loci of the group II enzymes, $P_{neut} = 0.43$.

Previously, Gillespie and Langley (1974) showed that the average heterozygosity ($\bar{H}_e$) per locus is much lower for substrate-specific than for the nonspecific enzymes not only for *Drosophila* but also for the human and the mouse. These observations are compatible with the neutral theory if it is assumed that selective constraint (negative selection) is stronger for substrate-specific than nonspecific enzymes. This means that the probability of an amino acid change being not harmful, that is, selectively neutral, is smaller for the substrate-specific than for nonspecific enzymes, even if the total mutation rate per locus ($v_{T(E)}$) is the same for these two types of loci.

Finally, I would like to show that the present analysis on $P_{neut}$ is consistent with the results obtained from recent studies on the evolutionary rate of globin pseudogenes. As shown by Miyata and Yasunaga (1981) and Li et al. (1981), the evolutionary rate of nucleotide substitutions is very high for pseudogenes. This is easily understandable from the neutral theory, because pseudogenes can be regarded as "dead genes" which have been liberated from the constraint of neg-

ative selection, so that all the mutations in them become selectively neutral. Thus, pseudogenes accumulate mutational changes at the maximum speed as predicted by the neutral theory. This can be explained in more quantitative terms as follows: if I denote the fraction of neutral mutations by $f_0$ (which is determined by the degree of selective constraint), the rate of evolution in terms of mutant substitutions is

$$k = v_0 = f_0 v_T, \tag{11}$$

where $v_0$ is the neutral mutation rate and $v_T$ is the total mutation rate. Note that, under the neutral theory, the rate of evolution is equal to the mutation rate for neutral alleles (Kimura 1968*a*). As predicted by Kimura (1977), the maximum evolutionary rate is attained when $f_0 = 1$, and it is likely that pseudogenes indeed represent such a case.

If I adopt the estimates given in table 3 of Li et al. (1981), the average rate for the three globin pseudogenes, mouse ψα3, human ψα1, and rabbit ψβ2, is $4.6 \times 10^{-9}$ substitutions per nucleotide site per year. However, the rates of nucleotide substitutions at the first, second, and third positions of the codons in the normal globin genes are $0.71 \times 10^{-9}$, $0.62 \times 10^{-9}$, and $2.64 \times 10^{-9}$, respectively (Li et al. 1981). In order to estimate the mutation rates $v_{T(E)}$ and $v_{0(E)}$ from these observed values, one needs to know what fraction of nucleotide changes at each of the three positions of the codon cause electrophoretically detectable amino acid changes. For this purpose, assume that electrophoretic mobility of a protein is determined solely by its net charge and that, among 20 amino acids, aspartic and glutamic acids are acidic and negatively charged, lysine and arginine are basic and positively charged, while the rest are electrically neutral. Then, from the standard code table, we find that the probability of a random nucleotide change causing an electrophoretically detectable amino acid change is about 0.28 for the first position, one-third for the second position, and only one-twelfth for the third position of the codon.

I also note that nucleotide changes always cause amino acid changes at the second position and predominantly so at the first position. However, at the third position, nucleotide changes cause amino acid changes in only some one-third of the cases, the rest being synonymous. Furthermore, in globins, the synonymous component of nucleotide substitutions has an evolutionary rate at least two or three times as high as the amino acid altering nucleotide substitutions (Jukes 1980; Kimura 1981), suggesting that the probability of a random nucleotide change being selectively neutral is much higher for the synonymous than for amino acid altering changes. There is also the phenomenon of nonrandom usage of synonymous codons (Grantham 1980; Ikemura 1981), and this, too, complicates the problem. For these reasons, I exclude the data from the codon's third position in the following calculation.

Then, using Li et al.'s (1981) estimates, I can compute the fraction of neutral mutations with respect to electrophoretically detectable changes in hemoglobin by the ratio $(0.71 \times 0.28 + 0.62/3)/(4.6 \times 0.28 + 4.6/3)$, which gives $P_{neut}(\text{Hb}) = 0.14$. Although I do not know the evolutionary rates in terms of amino acid substitutions of the various enzymes and other proteins used to estimate $P_{neut}$ in table 1, it is likely that their average evolutionary rate is not very different from the evolutionary rate of hemoglobin which is near the median of the evolutionary rates of proteins (Kimura 1974). Considering the many uncertainties involved in

the process of estimating the fraction of neutral mutations, the agreement between the two independent estimates above, that is, $P_{neut} = 0.14 \pm 0.06$ for enzyme and other protein loci in the four organisms and $P_{neut}(Hb) = 0.14$ for hemoglobin in mammals, is impressive. I believe that this consistency strongly supports the neutral theory. I also think that a detailed study of rare variant alleles is just as important for understanding the mechanism of the maintenance of genetic variability as that of polymorphic alleles. It is hoped that more data on rare variants will be obtained for wild species whose ecologies, particularly the population sizes, are well known.

**Acknowledgment**

This work is supported in part by a grant-in-aid from the Japanese Ministry of Education, Science and Culture.



LITERATURE CITED

AYALA, F. J., M. L. TRACEY, L. G. BARR, J. F. MCDONALD, and S. PÉREZ-SALS. 1974. Genetic variation in natural populations of five *Drosophila* species and the hypothesis of the selective neutrality of protein polymorphisms. Genetics **77:**343–384.

CALDER, N. 1973. The life game. BBC, London.

CHAKRABORTY, R. 1981. Expected number of rare alleles per locus in a sample and estimation of mutation rates. Amer. J. Human Genet. **33:**481–484.

CROW, J. F. 1972. The dilemma of nearly neutral mutations: how important are they for evolution and human welfare? J. Hered. **63:**306–316.

———. 1981. The neutralist-selectionist controversy: an overview. Pp. 3–14 *in* E. B. HOOK, ed. Population and biological aspects of human mutation. Academic Press, New York.

GILLESPIE, J. H., and C. H. LANGLEY. 1974. A general model to account for enzyme variation in natural populations. Genetics **76:**837–848.

GRANTHAM, R. 1980. Workings of the genetic code. Trends Biochem. Sci. **5:**327–331.

HARRIS, H. 1976. Molecular evolution: the neutralist-selectionist controversy. Fed. Proc. **35:**2079–2082.

HARRIS, H., and D. A. HOPKINSON. 1972. Average heterozygosity per locus in man: an estimate based on the incidence of enzyme polymorphisms. Ann. Human Genet. **36:**9–20.

HARRIS, H., D. A. HOPKINSON, and E. B. ROBSON. 1974. The incidence of rare alleles determining electrophoretic variants: data on 43 enzyme loci in man. Ann. Human Genet. **37:**237–253.

IKEMURA, T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNA and the occurrence of the respective codons in its protein genes. J. Mol. Biol. **146:**1–21.

JUKES, T. H. 1980. Silent nucleotide substitutions and the molecular evolutionary clock. Science **210:**973–978.

KIMURA, M. 1968*a*. Evolutionary rate at the molecular level. Nature **217:**624–626.

———. 1968*b*. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet. Res. **11:**247–269.

———. 1974. Gene pool of higher organisms as a product of evolution. Cold Spring Harbor Symp. Quant. Biol. **38:**515–524.

———. 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature **267:**275–276.

———. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Natl. Acad. Sci. USA **78:**454–458.

———. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

KIMURA, M., and J. F. CROW. 1964. The number of alleles that can be maintained in a finite population. Genetics **49:**725–738.

KIMURA, M., and T. OHTA. 1971. Protein polymorphism as a phase of molecular evolution. Nature **229:**467–469.

LEWONTIN, R. C. 1974. The genetic basis of evolutionary change. Columbia University Press, New York.

LI, W.-H., T. GOJOBORI, and M. NEI. 1981. Pseudogenes as a paradigm of neutral evolution. Nature **292:**237–239.

MIYATA, T., and T. YASUNAGA. 1981. Rapidly evolving mouse α-globin-related pseudogene and its evolutionary history. Proc. Natl. Acad. Sci. USA **78:**450–453.

NEEL, J. V. 1978. Rare variants, private polymorphisms, and locus heterozygosity in Amerindian populations. Amer. J. Human Genet. **30:**465–490.

NEI, M. 1975. Molecular population genetics and evolution. North-Holland, Amsterdam.

———. 1977. Estimation of mutation rate from rare protein variants. Amer. J. Human Genet. **29:**225–232.

NEI, M., R. CHAKRABORTY, and P. A. FUERST. 1976. Infinite allele model with varying mutation rate. Proc. Natl. Acad. Sci. USA **73:**4164–4168.

NEVO, E. 1978. Genetic variation in natural populations: patterns and theory. Theoret. Pop. Biol. **13:**121–177.

NOZAWA, K., T. SHOTAKE, Y. KAWAMOTO, and Y. TANABE. 1982. Population genetics of Japanese monkeys. II. Blood protein polymorphisms and population structure. Primates **23:**252–271.

NOZAWA, K., T. SHOTAKE, and Y. OKURA. 1975. Blood protein polymorphisms and population structure of the Japanese macaque, *Macaca fuscata fuscata*. Pp. 225–241 *in* C. L. MARKERT, ed. Isozymes IV: genetics and evolution. Academic Press, New York.

OHTA, T. 1975. Statistical analyses of *Drosophila* and human protein polymorphisms. Proc. Natl. Acad. Sci. USA **72:**3194–3196.

———. 1976. Role of very slightly deleterious mutations in molecular evolution and polymorphism. Theoret. Pop. Biol. **10:**254–275.

RUFFIÉ, J. 1976. De la biologie à la culture. Flammarion, Paris.

SELANDER, R. K. 1976. Genetic variation in natural populations. Pp. 21–45 *in* F. J. AYALA, ed. Molecular evolution. Sinauer, Sunderland, Mass.

WARD, R. D., and J. A. BEADMORE. 1977. Protein variation in the plaice, *Pleuronectes platessa* L. Genet. Res. **30:**45–62.

MASATOSHI NEI, reviewing editor

Received June 22, 1983; revision received August 2, 1983.